REMARKS

I. Claim Amendment

Claim 8 has been amended to replace the Roman numerals, by which the subsets of step (g) were numbered, to Arabic numerals.

Claim 34 has been amended to correct typographical error. No new matter has bben introduced by this amendment.

Claims 37 has been amended to recite that the digital processor provides the produced classification as output.

No new matter has been introduced by these amendments.

II. Claim Rejection Under 35 U.S.C. §102(e)

Claims 2, 3, 7-9, 11, 16, 22-25, 27-30, 34 and 37 are rejected under 35 U.S.C. §102(e) as anticipated by U.S. Pat. No. 6,834,239 to Lobanov *et al.*

Without recapitulating the arguments presented in the Office Action, Applicants disagree with the Examiner's characterization of U.S. Pat. No. 6,834,239 and maintain that U.S. Pat. No. 6,834,239 does not anticipate the claimed invention. Applicants reserve the right to present further arguments substantiating their position.

With the instant Reply, Applicants submit Declaration by Beth T. Logan and Pedro J. Moreno under 37 C.F.R. §1.131 (hereinafter, the Logan Declaration). The Logan Declaration antedates U.S. Pat. No. 6,834,239, which has the earliest effective prior art date under 35 U.S.C. §102(e) of August 22, 2000, prior to applicant's effective filing date of November 28, 2000. (MPEP 715 (I)(B))

According to 37 C.F.R. §1.131(b):

> The showing of facts [in a Declaration] shall be such, in character and weight, as to establish reduction to practice prior to the effective date of the reference, or conception of the invention prior to the effective date of the reference coupled with due diligence from prior to said date to a subsequent reduction to practice or to the filing of the application.

The Logan Declaration demonstrates both reduction of the present invention to practice by July 5, 2000 and conception of the present invention at least by late May 2000 coupled with due

diligence in reducing the present invention to practice from May 2000 to July 5, 2000 and further on, up to filing the instant application on November 28, 2000.

Specifically, the Logan Declaration presents evidence (Exhibits A - F) of an effort by Simon Kasif, Beth T. Logan, and Pedro J. Moreno, and Baris E. Suzek to develop a novel approach to protein classification whereby proteins would be represented by combining small sequences. Mr. Suzek, an intern at COMPAQ Computer Corp. during the Summer of 2000, maintained a logbook of his efforts on a source-controlled internal company website (hereinafter, the Website). Because the Website is source-controlled, it is possible to authenticate and date each of the website printouts, presented as Exhibits A through F of the Logan Declaration.

Accordingly, the Logan Declaration presents and attests to the following facts:

(1)   A statement by the signatories that the conception of protein classification using a combination of short sequences contained within the protein sequence happened as early as late May 2000. (The Logan Declaration, para. 3.)

(2)   An entry by Mr. Suzek on the Website dated June 9, 2000 which recorded that the method under design would find short sequences within the protein sequence being analyzed that have the best match with an existing database. (The Logan Declaration, para. 4 and Exhibit A.)

(3)   An entry by Mr. Suzek on the Website dated June 12, 2000 that short protein sequences from an existing and publicly accessible database BLOCKS were being used for both development and validation of the method under design. (The Logan Declaration, para. 5 and Exhibit B.)

(4)   An entry by Mr. Suzek on the Website dated June 13, 2000 which recorded that protein sequences were being modeled as concatenation of short segments and that an existing and publicly accessible database SCOP was being used to validate the method under development. (The Logan Declaration, para. 6 and Exhibit C.)

(5)   A statement by the signatories that, by June 19, 2000, the idea of scoring the input protein sequences by counting the number of times each short segment is contained within the sequence had been implemented. (The Logan Declaration, para. 7, and Exhibit D.)

(6)   An entry by Mr. Suzek on the Website dated June 20, 2000 which recorded that the feature vectors generated to describe the input protein sequences can be classified using any

of a number of multidimensional vector classification techniques such as Singular Value Decomposition and Support Vector Machines. (The Logan Declaration, para. 8 and Exhibit D.)

(7)   An entry by Mr. Suzek on the Website dated July 7, 2000 which recorded ideas put forth during the meeting that took place on June 27, 2000 regarding the refinement of the methods of scoring an input protein sequence against the database of short segments. (The Logan Declaration, para. 10 and Exhibit F.)

(8)   An entry by Mr. Suzek on the Website dated July 7, 2000 which recorded completion by July 5, 2000 of a fully operational embodiment of the method of the present invention that employs Support Vector Machines method of vector classification. (The Logan Declaration, para. 11 and Exhibit F.)

(9)   A presentation given by Mr. Suzek in early August 2000 that describes one of the embodiments of the present invention. (The Logan Declaration, para. 12 and Exhibit G; Exhibit H (Invention Disclosure), para. 3E.)

The Logan Declaration further presents the following exhibits documenting due diligence:

(10)   A copy of an electronic communication dated August 21, 2000 from Ms. Logan to Mr. R. Reed, an engineering liaison to a legal counsel for COMPAQ Computer Copr., forwarding the completed draft of the invention disclosure. (The Logan Declaration, para. 13 and Exhibit H.)

(11)   A copy of an electronic communication dated September 15, 2000 from Mr. R. Lange, a Legal Counsel for Compaq Computer Corp., contacting Ms. MaryLou Wakimura, a principal at the law firm of Hamilton, Brook, Smith & Reynolds with a request to prepare the patent application based on the research work described above. (The Logan Declaration, para. 14 and Exhibit I.)

(12)   A copy of an electronic communication dated September 22, 2000 from Ms. Logan to Ms. Wakimura scheduling a meeting between the inventors and Ms. Wakimura that took place on October 5, 2000. (The Logan Declaration, para. 15 and Exhibit J.)

Applicants submit that the facts presented and evidenced by the Logan Declaration demonstrate that the present invention was actually reduced to practice by July 5, 2000, almost two months prior to the earliest effective priority date of U.S. Pat. No. 6,834,239. Applicants

also submit that the facts and the evidence demonstrate that the invention was conceived at least in late May 2000 and that the conception was followed by due diligence in reducing the invention to practice. Applicants further submit that the Logan Declaration satisfies the burden of proof under 37 C.F.R. §1.131.

Thus, the rejection under 35 U.S.C. §102(e) cannot stand where the cited reference, U.S. Pat. No. 6,834,239, was not "... filed in the U.S. before the invention by the applicant." (*Emphasis added.*)

Reconsideration and withdrawal of the rejection are respectfully requested.

II. Claim Rejection Under 35 U.S.C. §112, First Paragraph

(1)   *Summary of the Invention*

The present invention is a method and an apparatus for analyzing a sequence from a subject's genome by assigning this sequence to a class. The method operates by accepting two inputs: (1) a set of "known biological fragments", *i.e.* characteristic biological sequences obtainable from available databases, *e.g.* protein domains; and (2) a genome sequence to be analyzed.

During the operation of the method of the invention, a "feature vector" is created for each genome sequence that is being analyzed. A "feature vector" represents the genome sequence in a manner that is readily accepted by a programmable machine. A "feature vector" is a list of numerical scores of each fragment from the set of "known biological fragments". The numerical scores are obtained by measuring the likelihood of finding a given fragment within the subject genome sequence that is being analyzed. Advantageously, the use of these "feature vectors" allows one to compare genome sequences of variable length since each "feature vector" always has the same number of elements equal to the number of fragments in the set of "known biological fragments". This feature of the method is referred to as "uniform" representation.

Implicit in any process of assigning an object to a class is comparison between the object being assigned and the objects that had already been assigned to this class. In the context of the present invention, the genome sequence being analyzed is referred to as a "query" (specifications page 11, line 5), while the sequences that had already been classified are referred to as a "training set" (specification page 10, lines 18-20).

One skilled in the art will immediately appreciate that a training set of sequences can either be provided (specification page 10, lines 18-25 and Example, pages 11-13), or, when more than one subject's genome sequences are being analyzed, the earlier analyzed sequences can be considered a "training set" with respect to later analyzed sequences. However, irrespective of the source, the training sequences must be converted into feature vectors (page 11, lines 14-19) to afford application of the classifying techniques. Thus, a set of training sequences is necessarily converted into a set of training vectors. It is the training vectors that are being classified by a programmable machine. See specification page 11, lines 1-12 (describing "N12" indexing technique); pages 10-11 (describing k-mean clustering and nearest neighbor analysis); specification page 10, lines 1-2 (describing principal component analysis (PCA)); and pages 11-13 (describing Support Vector Machines).

(2)   *Applicants' Arguments*

Claims 2, 3, 7-9, 11, 16, 22-25, 27-30, 34 and 37 are rejected under 35 U.S.C. §112, first paragraph, as containing subject matter not described in the specification in such a way as to convey to one skilled in the art that the inventors had possession of the claimed invention at the time of filing. The Examiner further stated that this is a new matter rejection.

(A) *Claim 34*

Specifically, the Examiner stated that Claim 34, from which Claims 2, 3, 7-9, and 22-25 depend, as well as Claim 37 describe methods and apparatus not *specifically* or *concretely* disclosed in the specification as filed and are therefore new matter.

Applicants respectfully bring into the Examiner's attention that "specific" or "concrete" disclosure is not a requirement under 35 U.S.C. §112 or the applicable Federal Rules. It is Applicants understanding that the Examiner is asserting the lack of literal or express disclosure supporting Claims 34 and 37.

It is well-established in the Patent Law, that there is no *in haec verba* requirement and that newly added claim limitations must be supported in the specification through either express, implicit, or inherent disclosure. (See MPEP 2163.I.B, citing *In re Oda*, 443 F.2d 1200, 170 USPQ 268 (CCPA 1971).) It is further well-established, that an amendment to correct an obvious error does not constitute new matter where one skilled in the art would not only

recognize the existence of the error in the specification, but also recognize the appropriate correction. (*Ibid.*) Furthermore, argument of lack of literal support is insufficient to support a rejection under 35 U.S.C. §112, first paragraph. In re Wertheim, 541 F.2d 257, 262, 191 USPQ 90, 96 (CCPA 1976).

Furthermore, in BEHR ET AL. v. TALBOTT ET AL., Patent Interference No. 102,366, Board of Patent Appeals and Interferences of the Patent and Trademark Office (July 1, 1992; Final Hearing March 10, 1992), the Board explained:

> The subject matter of the claim need not be described identically or literally for the application to satisfy the written description requirement. Kennecott Corp. v. Kyocera Int'l Inc., 835 F.2d 1419, 1422, 5 USPQ2d 1194, 1197 (Fed.Cir.1987), cert. denied, 108 S.Ct. 1735 (1988); In re Wilder, 736 F.2d 1516, 1520, 222 USPQ 369, 372 (Fed.Cir.1984). [...]

> The written description requirement can be satisfied by showing that the disclosed subject matter, when given its "necessary and only reasonable construction," inherently (i.e., necessarily) satisfies the limitation in question. Kennecott v. Kyocera, 835 F.2d at 1423, 5 USPQ2d at 1198 ("It was conceded that anyone with a microscope would see the [equiaxed] microstructure of the product of the '954 application"); Snitzer v. Etzel, 531 F.2d at 1076, 189 USPQ at 419; Wagoner v. Barger, 463 F.2d at 1380, 175 USPQ at 86-87; Binstead v. Littmann, 242 F.2d at 770, 113 USPQ at 281-82.

The Examiner stated that the examples in the specification do not set forth a method having each step (a)-(i) of Claim 34 nor is it clear from the specification that this particular method was originally disclosed or envisioned.

With the reference to the *Summary of the Invention* section of the instant Amendment, Applicants submit that steps (a) through (e) of Claim 34 amount to a recitation of the steps required to create a set of training vectors; steps (f) - (i) amount to classifying an input genome sequence as claimed by the original Claims 1 and 5 as filed.

Creation of a set of training vectors is both expressly described (see pages 10-13 of the specification as filed) and implicitly required for an operation of a method that attempts to classify vectors. Indeed, step (a) of Claim 34 recites:

> (a) providing a set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation;

Applicants direct the Examiner's attention to page 8, line 9 through page 9, line 2 of the specification as filed. The indicated portion of the specification describes that in order to create feature vectors, the invention method first obtains a comparison database of short, highly conserved regions in related protein domains, *i.e.* fragments also referred to as "blocks", "motifs" or "probabilistic templates". The specification further describes that any such fragment is preferably represented (*i.e.* has a respective representation) by a K by L matrix M in which each of the K rows represents a particular amino acid (or nucleotide for DNA sequences) and L represents the length of the motif. Applicant submit that this portion of the specification as filed provides express support for step (a) of Claim 34.

Step (b) of Claim 34 recites:

(b)   providing at least one training sequence;

Applicants direct the Examiner's attention to page 10, lines 18-25 and page 11, lines 16-19 of the specification as filed. The indicated portions of the specification as filed describe that the process of classification attempts to create a model based on a set of labeled feature vectors called the "training set". Given a set of training protein sequences labeled according to structure, the method of the invention converts each of these sequences into a multidimensional feature vector (*i.e.* the method converts a set of training sequences into a set of training vectors). Applicants submit that this description necessarily implies that at least one training sequence is provided.

Step (c) of Claim 34 recites:

(c)   for each known biological fragment, quantitatively determining a score with respect to each training sequence;

Applicants submit that since the method of the present invention operates to classify biological sequences by converting them into vectors, it follows that any sequence being analyzed, including a training sequence, is converted into a feature vector. The feature vector, as described throughout the specification and in the above-quoted portion of the specification and in particular on page 9, lines 4 to 6 of the specification as filed, is created by quantitatively determining a score of each of the known biological fragment, with respect to each biological sequence.

Step (d) of Claim 34 recites:

(d)   for each training sequence, forming a training feature vector, said training feature vector being a sequence of scores of each known biological fragment with respect to the training sequence;

Applicants direct the Examiner's attention to page 11, lines 16-19 of the specification as filed, which expressly describe that training sequences are converted into training vectors.

Step (e) of Claim 34 describes creation of classes using the training set of vectors. With the reference to the *Summary of the Invention* section of the instant Amendment, Applicants submit that this step is both implicitly required for classification of an input genome sequences and is expressly described in the specification as filed. Indeed, Applicants direct the Examiner's attention to page 11, lines 20-24:

2.   Given the labeled feature vectors generated in step 1 [*i.e. a set of training vectors*], learn corresponding Support Vector Machine (SVM) classifiers (Burger, 1998, "A tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery Journal*) to separate each structural class from "the rest of the world". A SVM classifier learns a separating hyperplane between two classes which maximizes the "margin"--the distance between the hyperplane and the nearest datapoint of each class.

Furthermore, one skilled in the art will understand that whether the classification algorithm is a SVM or any of the algorithms described in the *Summary of the Invention* section of the instant Amendment, such algorithms are necessarily trained using the training set. The specific classification algorithm is not a part of the inventive concept and should not be read into the claim.

With respect to steps (f), (g), (h) and (i), Applicants submit that these steps recapitulate the original Claim 1 and 5 as filed. Support for these steps is found throughout the specification.

Thus, each and every step of Claim 34 is supported, expressly or implicitly, by the instant disclosure as originally filed. No new matter was introduced by this claim. Reconsideration and withdrawal of the rejection are respectfully requested.

(B) *Claim 3*

The Examiner stated that Claim 3 as amended does not have the basis in the specification. It is Applicants' understanding that the Examiner is objecting to the term "plurality" which refers

to genome sequences used in step (f) of Claim 34. Examiner's attention is directed to the original Claim 12 as filed, which recited "a plurality of different subject genome sequences; and wherein the comparison routine forms for each subject genome sequence, a respective vector such that a corresponding plurality of same length vector representations is provided." Applicants submit that Claim 3 recapitulates the original Claim 12 as filed.

No new matter was introduced by this claim. Reconsideration and withdrawal of the rejection are respectfully requested.

(C)   *Claim 8*

The Examiner stated that Claim 8 as amended does not appear to have basis in the originally filed specification.

Applicants note that Claim 8 as amended recites a procedure by which step (g) of Claim 34 operates. In step (g) of Claim 34, a score of each known biological fragment with respect to the subject genome sequence is quantitatively determined. As described on page 4, line 20-26, as well as throughout page 9 of the specification as filed, each individual score (one for each known sequence in the database) serves as a feature vector element.

Applicants direct the Examiner's attention to Claims 1 and 8 as originally filed. The original Claim 1 recited that scoring each biological fragment against the genomic sequence includes "(i) counting the number of times the known biological fragment is found in the subject genome sequence and (ii) from said counted number of times, forming a vector element". The original Claim 8, dependent on Claim 1, recited that "counting includes determining probability of the subject genome sequence being generated by the known biological fragment." Applicants submit that Claim 8 as amended merely recites subject matter of the original Claim 8. Furthermore, Applicants submit that Claim 8 as amended is further expressly supported on page 9, lines 3-9 of the specification as filed.

No new matter was introduced by this claim. Reconsideration and withdrawal of the rejection are respectfully requested.

(C) *Claims 22-25*

It is Applicants' understanding that Claims 22-25 are objected to as dependent on a rejected Claim 34.

Applicants refer to the above-presented arguments with respect to Claim 34.

Reconsideration and withdrawal of the rejection are respectfully requested.

(D) *Claim 37*

The Examiner stated that Claim 37 lacks specific disclosure in the original application and lacks means for outputting the results.

Applicants submit that Claim 37, from which Claims 11, 16 and 27-30 depend, is a claim to an apparatus that performs the steps of the method claimed in Claim 34. Claim 37 incorporates the steps of the original Claims 10, 12 and 13 as filed. Step (4) of Claim 37 recites the steps (e) and (i) of Claim 34. Claim 37 has further been amended to include output means. With reference to the MPEP 2163.I.B, citing *In re Oda*, 443 F.2d 1200, 170 USPQ 268 (CCPA 1971), Applicants submit that an omission of output means in an apparatus having input means, a memory (data store) and a digital processor is an obvious error.

(E) *Claims 11, 16 and 27-30*

It is Applicants' understanding that Claims 11, 16 and 27-30 are objected to as dependent on a rejected Claim 37.

Applicants refer to the above-presented arguments with respect to Claim 37.

Reconsideration and withdrawal of the rejection are respectfully requested.

III. Claim Rejection Under 35 U.S.C. §112, Second Paragraph

Claims 2, 3, 7-9, 11, 16, 22-25, 27-30, 34 and 37 are being under 35 U.S.C. §112, second paragraph, rejected as indefinite. It is Applicants' understanding that Claims 34 is being objected to as containing a typographical error and that Claim 37 is being rejected as failing to recite a component of the claimed apparatus. It is further Applicants' understanding that the remainder of the claims is objected to as being dependent on the rejected/objected claims. If Applicants are incorrect, clarification is respectfully requested.

With respect to Claim 34, Applicants amended this claim to correct the typographical error. The step following step (h) is now labeled (i). Applicants note that Claim 8 does NOT refer to step (i) of Claim 34, but rather recites two subsets of step (g) of Claim 34. To make the enumeration within Claim 8 clear, Claim 8 is now amended to recite subsets of step (g) as "(a)" and "(2)". The terms "subset 1" and "subset 2" is not a reference to step (i) of base Claim 34.

With respect to Claim 3, in order to be consistent with making genome sequences plural in step (f), Claim 3 recites that step (h) of Claim 34 handles each one of the plurality. To further address the Examiner's stated concern, the preamble of Claim 34 is now amended to be directed to classifying *one or more* genomic sequences.

With respect to Claim 37, Applicant submit that this claim cannot be deemed defective. MPEP 2106.IV.B.2.(a) clearly explains:

> If a claim defines a useful machine or manufacture by identifying the physical structure of the machine or manufacture in terms of its hardware or hardware and software combination, it defines a statutory product. See, e.g., *Lowry*, 32 F.3d at 1583, 32 USPQ2d at 1034-35; *Warmerdam*, 33 F.3d at 1361-62, 31 USPQ2d at 1760.

Claim 37 defines a machine comprising an input device, a data store and a digital processor. It is distinguishable from prior art by the configuration of the processor, which determines how the machine functions. Furthermore, MPEP 2106.V.A.2 states:

> However, the applicant need not explicitly recite in the claims every feature of the invention. For example, if an applicant indicates that the invention is a particular computer, the claims do not have to recite every element or feature of the computer. In fact, it is preferable for claims to be drafted in a form that emphasizes what the applicant has invented (i.e., what is new rather than old). *In re Dossel*, 115 F.3d 942, 946, 42 USPQ2d 1881, 1884 (Fed. Cir. 1997).

Applicants submit that rejection of Claim 37 is improper.

With respect to Claims 11, 16, 27 and 29, Applicants request clarification of the nature of the rejection. Examiner stated that Claims 11, 16, 27 and 29 are drawn to non-functional descriptive material and do not further limit the structure of the apparatus itself. MPEP 2106.IV.B.1.(b) states with respect to non-functional descriptive material:

> Descriptive material that cannot exhibit any functional interrelationship with the way in which computing processes are performed does not constitute a statutory process, machine, manufacture or composition of matter and should be rejected under 35 U.S.C. 101.

Applicants request clarification as to whether Claims 11, 16, 27 and 29 are rejected under 35 U.S.C. §101. Applicants further note as Claims 11, 16, 27 and 29 depend on Claim 37, not

rejected under 35 U.S.C. §101. Accordingly, Claims 11, 16, 27 and 29 cannot be rejected on this basis.

The preceding arguments notwithstanding, Applicants submit that Claims 11, 16, 27 and 29 are *not* drawn to non-functional descriptive material. Indeed, Claim 11 adds to Claim 37 a requirement that the data store be a published (*i.e.* publicly available) database of protein motives. Claim 16 introduces and limits the types of genomic sequences used as input. Namely, an input genome sequence can be a DNA or a protein sequence. Claim 27 requires that each biological fragment be stored as a text string (as opposed to a probabilistic template), while Claim 29 requires the opposite: storage of each fragment in the form of a probabilistic template rather than a text string.

Since the elements recited by Claims 11, 16, 27 and 28 exhibit functional interrelationship with the way in which computing processes are performed, they cannot be deemed non-functional descriptive material.

Reconsideration and withdrawal of the rejection are respectfully requested.

CONCLUSION

In view of the above amendments and remarks, it is believed that all claims are in condition for allowance, and it is respectfully requested that the application be passed to issue. If the Examiner feels that a telephone conference would expedite prosecution of this case, the Examiner is invited to call the undersigned.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By _____

Alexander Akhiezer
Registration No. 54,617
Telephone: (978) 341-0036
Facsimile: (978) 341-0136

Concord, MA 01742-9133
Dated: 8/8/05